DOCUMENT RESUME

ED 126 880 IR 003 793

AUTHOR Mielke, Keith W.
TITLE Research and Evaluation in Educational Television.
INSTITUTION Stanford Univ., Calif. School of Education.
SPONS AGENCY Ford Foundation, New York, N.Y.
PUB DATE May 73
NOTE 49p.; Chapter 8 from "Educational Television: A Policy Critique and Guide for Developing Countries"

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Administrator Evaluation; Communications; Decision Making; *Educational Television; *Evaluation; Evaluation Methods; Formative Evaluation; Media Research; Performance Criteria; Problem Solving; Program Evaluation; Research; *Research Criteria; Research Methodology; Summative Evaluation; *Television Research
IDENTIFIERS Administrative Research

ABSTRACT

A fundamental distinction is made between basic research which seeks out new, generalizable knowledge, and administrative research which provides information useful to decision makers who have a task to perform. Attention is then given to four types of administrative research: background research, formative research, summative research, and policy research. A concluding series of questions and comments about a hypothetical educational television program puts the separate research branches back into their natural interrelationships. (CH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED126880

RESEARCH AND EVALUATION IN EDUCATIONAL TELEVISION

Categories: The Imperfect but Indispensable Tools

A story that has served long and well to illustrate a basic point in semantics goes something like this: Three baseball umpires were discussing how they reach decisions when judging baseball games. The first umpire said, "Some are strikes; some are balls; and I call them as they are." The second umpire said, "Some are strikes; some are balls; and I call them as I see them." The third umpire said, "Some may be called strikes and some may be called balls, but they aren't anything until I call them."

In talking about research/evaluation instead of baseball, the different views of reality represented by the umpires in the story are still with us. The first umpire would probably say that some project is or is not research, as if he were making a simple factual statement. The second umpire holds the same basic beliefs as the first, except he recognizes that his observational powers are not infallible. It is only umpire three who appreciates fully the creative power of the label or category. Given the variety of research labels now in use throughout the literature, the research categories used in this chapter should be viewed through the eyes of umpire number three; not as having the quality of correctness, but, hopefully, as having the quality of utility.

Overview of the Chapter

First, a fundamental distinction is made between basic research (seeking new, generalizable knowledge) and administrative research

IR003793

providing information useful to decision-makers who have a mission to perform). Attention is then given to four types of administrative research: background research, formative research, summative research, and policy research. Having thus subdivided a complex process into analytic categories, a concluding series of questions and comments about a hypothetical ETV program will put the separate parts back into their natural inter-relationships.

The disproportionate emphasis on administrative research does reflect a value judgment, not on the superiority of one research method or philosophy over another, but on what will be most relevant to the needs of an ETV administrator. Administrative research should help people make decisions and solve problems. The various categories of research discussed here, therefore, should serve as an inventory of ways that research can be useful.

Basic Research

In general, ETV organizations are not directly involved in basic communication research; therefore the bulk of this chapter will not deal directly with this category. Basic research is, however, a useful point of reference for describing, contrasting, and understanding the administratively oriented research more typical of ETV operations, so it will be discussed first.

Basic research is sometimes called "pure" research or theoretical research. A distinguishing characteristic is its primary purpose: increasing generalizable knowledge and understanding of basic principles through the construction, testing, and validation of theories (explanations of relationships). Predictions are derived logically from

the theories for empirical tests in a research design. The major research criteria are internal validity (Can the results be interpreted unambiguously?), external validity (Are the findings generalizable?), and theoretic relevance (Was the right question asked in the first place? Was the theoretic explanation actually tested in this set of observations?).

In basic research, stimulus materials would be valued, not for what they <u>are</u>, but for the theoretical conditions that they <u>represent</u>. Suppose a basic researcher is studying televised violence and its effects. As part of his experimental requirements, he uses a film clip of a brutal fight scene. Does this researcher have any interest in this particular fight scene per se? No; it was chosen from any number of alternative film clips that could represent the needed degree of violence. The researcher's interest would be in getting the required levels of violence clearly and efficiently represented so that his experimental results can be unambiguous. The particular fight scene employed, and/or the movie from which it was taken, is of little or no ultimate interest to the basic researcher. Later, it will be seen that this contrasts markedly with the administrative researcher, whose interest <u>will</u> be in a particular program as a program, not as a replaceable representative of some theoretic condition.

Basic research thrives in a <u>disciplinary</u> environment, where it serves to develop the discipline. Minimal conditions needed to establish a discipline would appear to be: (1) a cohesive set of variables with which the discipline will be concerned; (2) an accepted research methodology or set of methodologies used to explore relationships among those variables; and (3) a body of theoretical knowledge that describes

and explains in generalizable terms the relationships among variables, as established by the disciplinary research. The cutting edge of disciplinary growth and change is basic/"pure"/ theoretical research.

Many disciplinary underpinnings for ETV can be found: e.g., psychology, social psychology, sociology, political science, economics, etc., plus a host of creative and performing arts. I have urged elsewhere (Mielke, 1972) a sensitivity to contributions from a wide variety of established disciplines, but not a total reliance on them for the basic research to devleop the theories and principles needed in applied communication organizations such as ETV. Media principles and practices, especially production practices, seem to be of only peripheral interest to most of the non-applied contributing disciplines. Were it not for the still-emerging field of communication research, this would leave unexamined a host of basic research questions that seek to probe scientifically the nature and potential of media such as television. Representative of the modest but growing body of basic communication literature of research and theory that deals explicitly with media characteristics is the work of Gavriel Salomon (e.g., Salomon and Snow, 1968; Salomon, 1970, 1972).

Rather than engaging primarily in basic research, ETV organizations typically have some social action mission assigned to them. Sometimes the ETV programmer can find guidance for his day-to-day operations from basic theory and research; sometimes not. If a basic researcher is studying how people of various ages and aptitudes learn to read, and if the ETV programmer has been assigned the mission of reducing illiteracy, then the theoretical insights of the basic researcher should be of great practical value to the ETV programmer.

In other situations, it may be far more difficult to determine the direct implications for ETV policy and practice from basic research.

Administrative Research

Techniques of inquiry can be and are used for purposes other than basic research, theory construction, and disciplinary development. Such techniques are also means of generating the data and information required for rational decision-making in administrative structures such as ETV organizations. This entire category of research designed to assist decision-making is being called administrative research.

There are some general characteristics that can distinguish administrative research from the basic research discussed earlier. The basic researcher sets his own constraints (the limits of his problem, the techniques to be employed, the factors to study, the factors to ignore, etc.), for the most part. The administrative researcher generally starts his work with most of these constraints assigned to him. Decision-making requirements dictate the problem and set constraints on what can be an acceptable answer.

The basic researcher deals with some form of the question: What is the effect of X on Y? Contrast this with the ETV administrative researcher's typical questions: Is there a need for this program? Is there an audience for this program? Will the target audience pay attention to this program? Will they comprehend it? Is one replay enough? Should we continue the program next year? Has the program met its objectives?

Guttentag (1971) writes: "Evaluation research always involves a judgment of the worthwhileness of some activity. At the outset,

therefore, it is quite different from the explicit value-free position of experimental research" (p. 76). (Note that Guttentag's statement does not imply a value-free position of the experimental researchers themselves.) In some form, administrative decision-making, and the administrative research that backs it up, deals with "the best way" (a value position) to allocate resources in order to meet some objective (with the objective being an accepted value position). All components of the organization, research included, are manifestations of a general commitment to find the "best" way to do a "good" thing.

Administrative research (sometimes called evaluation) deals with decision-making. There are also decisions to be made in basic research, of course, but they are formalized in advance to conform with scientific procedures for hypothesis acceptance or rejection. Ideally, in theory construction and disciplinary development, basic researchers should have the ability, even the responsibility, to hold commitments and beliefs in abeyance until there is sufficient supportive experimental evidence. Typically, the administrative researcher will not have the option to hold decision-making in abeyance. There is usually a finite set of options, with the decision-maker having to choose from that set on the basis of all the evidence, logic, intuitions, experience, advice, and value judgments at his command.

In basic research, there is more or less a <u>sole reliance</u> on evidence that has met the disciplinary requirements of acceptability. In administrative decision-making, the research evidence is one input alongside several others to be considered in the decision-making process.

In such "mother" disciplines as psychology, the controlled experiment is regarded highly, and rightfully so (see, e.g., Campbell,

1957; Kerlinger, 1965). To reduce extraneous influences, the basic researcher will devise stringent controls that attempt to hold all things equal or accounted for in two or more groups -- except some carefully manipulated factor. That "factor" will be a theoretic variable that is expected to make a difference. If a difference does occur, the good design will allow one to attribute the difference to the manipulated independent variable, because alternative explanations of the difference will have been made unlikely through careful controls. Although an oversimplified summary statement of experimental logic, this does allow us to contrast the administrative research situation. Through the eyes of a basic researcher in psychology, the ETV administrative researcher would not have sufficient control over the stimulus (the TV program), the subjects (the audience or the students), the design (the conditions of exposure), or even the dependent variable measurements (the criterion tests). The introduction of ETV usually means a whole new system for doing things. Not one, but many new factors or variables are involved. These factors are confounded and interrelated, and they make it very difficult to apply experimental logic. To a basic researcher, requirements of a rigorous design would override in importance the basic mission of a TV program. To an administrative researcher in ETV, maintenance of the basic mission of an ongoing program or series would override the demands of a rigorous experimental design.

All research, basic or administrative, costs money, but the conscious consideration of cost vs. payoff in research seems to be greater in administrative research. Conceivably, almost any reasonable administrative datum can be acquired for a price. The value judgment the administrator must make is how much of the system's resources can

be expended to find out something. This hits home immediately in ETV when, for example, a $10,000 program might cost half that much to evaluate properly. Therein lies one reason (but not the only reason) for a paucity of true experimental conditions being met in ETV administrative research. Sometimes administrative research is criticized for methodological compromises as if the researchers were unaware of what more ideal conditions might be, when in fact the compromise was dictated by expediency and economics. There is a considerable difference between the researcher who compromises in ignorance and the researcher who knows good research procedures and design, but must reluctantly deviate from the ideal in order to accomplish a mission.

For purposes of administrative research in ETV, classical experiments fail on several counts. Consider the following:

1. In exerting experimental control over the stimulus (the TV program), it may be simplified to the point where it no longer represents real world situations. Experiments that utilize visuals consisting of stick figures or simple line drawings, for example, are suspect in this regard.

2. In an experimental design set up to measure effects of, say, an entire series of ETV programs, other factors such as reception conditions, post-viewing reinforcement, etc., would have to be held constant throughout the series. If changed in midstream, it would be difficult to separate program effects from effects of the other sources of variance. Typically, an administrator would not tolerate this holding pattern if he saw something obviously going wrong. He would feel duty bound to improve his program as soon as possible and as much as possible. The experiment, per se, could yield no interim feedback for purposes of

corrective change. The experiment, then, can yield an after-the-fact judgment or "post mortem" on the series, but this is probably not the primary information need of the administrator.

3. In some experimental designs, the control group requirement means that a certain group should receive no treatment. This can be an ethical dilemma for the administrator who, almost by definition, believes in the value of his program. If the program is to bring good and needed results, should a group be deprived of exposure just to permit a clean-cut experimental assessment of effects? In ETV as well as in medicine and crime prevention, etc., the administrator's understandable response would often have to be "no."

4. Repeatedly, the requirements of tight design are impossible to implement in ongoing programs. For instance, subjects should be drawn from a target population and assigned to experimental treatment groups at random. This is usually impossible in situations where the researcher must work at the convenience of the education program, rather than vice versa.

5. The very rigor that gives experimental design its power also serves to shut out information that comes in forms outside the scope of the experiment. That is, experimental conclusions will be on the basis of quantitative measurement of the dependent variable. Any insights that are not reflected in that measurement have no effect on the conclusion.

At least on the surface, then, there seems to be a great amount of incompatibility between the values of basic research and the values of administrative research. Indeed, several have stated in some fashion that good basic research tends to make bad administrative research

(Guttentag, 1971; Stufflebeam, 1967; Guba, 1969). The position taken here is conciliatory. In particular, the values of theory and the logic of rigorous scientific research should not be forgotten, ignored, or disdained by the administrative researcher. Kurt Lewin's proposition that "there is nothing so practical as a good theory" still holds true for the administrative researcher, as does the desirability of unambiguous results. If Program A is working and Program B is not working, it is administratively necessary to be aware of that fact. Even more helpful in terms of future policy and practice, however, would be a compelling explanation of why that was so. Such is the value of theory. Even better would be such an explanation backed up with unequivocal empirical support. Such is the value of theory and scientific research working together.

Instead of the term "administrative research" used here, Stufflebeam, et al. (1971) use "evaluation," which they define as "the process of delineating, obtaining, and providing useful information for judging decision alternatives" (xxv). The authors make a succinct statement on the underlying reasons for similarities or differences between research and evaluation, and their statement serves as a fitting conclusion to this section:

> The purpose of research is to provide new knowledge, and its methodology is designed to produce knowledge that is universally valid. The purpose of evaluation, however, is to delineate, obtain, and provide information for making educational decisions. This information is not necessarily new knowledge, and even more important, it is highly particularistic and specific to a decision situation, rather than generalizable to many or all settings. Thus, evaluation methodology is not necessarily designed to produce universally valid information, but information that is valid and useful within the decision-making context. Insofar as the decision-making context is highly generalizable and the intent is to provide

> new information without precedent, the purposes and methodologies of research and evaluation may be equated (p. 140; emphases in the original).

Types of Administrative Research

With the generic categories of basic research and administrative research introduced and contrasted above, the exposition continues now to subdivisions of administrative research. For a variety of situations, consider what decision it is that must be made, who must make it, what information or data they need to make it, and what general principles or rules of decision-making they can, do, or should apply in reaching a decision. One can reason through such a set of questions and arrive at his own set of categories for administrative research. Chances are that the set would look something like the four categories below. More or fewer subdivisions could be made (recall the opening anecdote about the three baseball umpires) but the following categories allow distinctions to be made that are believed to be important:

A. Background research

B. Formative research

C. Summative research

D. Policy research.

Background Research

Something is needed to distinguish an important category of administrative research that is or can be independent of a particular TV program, product, or mission. It is called "background research" here; it could also be called "planning research" or "contextual research." For ETV, an important objective of background research would be general knowledge about the real and potential audience.

Consider such questions as the following:

1. What is the coverage area of our usable TV signal?
2. How many working receivers are contained within the signal area? Where are they located? Who has access to them?
3. What is the demographic composition of the total population, the potential TV audience, and the typical TV audience?

The three questions above typify a host of considerations that are indispensable in laying out general ETV programming strategies rationally.

In highly industrialized societies, a wealth of general reference data is supplied through business and government agencies. There are also specialized data services for broadcasters and advertisers. Even so, special background research studies must frequently be commissioned. With a decrease in access to standard reference data, there will be an increase in the need to conduct original background research. Through data on such questions as the following, background research can be helpful in suggesting programming needs:

Within various demographic sub-categories . . . .

4. . . . . to what interpersonal and mass media sources do people turn for advice on topics A, B, and C?

5. . . . . what are the areas of greatest concern?

6. . . . . what do people perceive as their greatest information needs?

Utilizing such questions, Mendelsohn (1968, 1969, 1971) has pioneered in using both demographic and psychological data as feedback to the

planning process in the area of social amelioration programming for television. In his Denver project (1968), for example, he determined through audience surveys that many mothers had sole responsibility for their households, were worried about money and health problems, and didn't know how or where to find assistance. These results provided guidelines to the writers and producers. Mendelsohn sometimes found that the facts gathered through such empirical background research contradicted the estimates of "expert" consultants.

Background research is also needed after a general programming strategy has been determined. Suppose, for whatever reasons, that the administrator is committed to produce an ETV series on agrarian reform, but no specifics as to scope or style of presentation have yet been decided. Consider the administrative utility of answers to the following questions, for example:

Within various demographic sub-categories . . .

7. . . . . what is presently known about modern agrarian practices? What general misconceptions are apparent? What are the largest gaps of knowledge?

8. . . . . what attitudes are held toward agrarian reform?

The ETV programmer must reach the target audience member on his (the target audience member's) terms, and objective background research data can illuminate the dimensions of this task. Consider, for example, the programming implications for almost any social amelioration ETV series, given a situation such as that found by the Paraguayan researcher Juan Dize Bodenave (1965). Bodenave interviewed 221 peasant farmers who lived about two hours' distance from Recife, Brazil. At the time of the study (1963), Recife, a state capitol, had almost a million inhabitants.

His study revealed that . . .

> . . . 94% of the peasants did not know that coffee was the principal export product of Brazil;
> . . . 80% had no definition for the word "democracy";
> . . . 73% had not heard of Fidel Castro;
> . . . 65% had no idea of what a dollar was;
> . . . 48% had not heard the name of the President of Brazil.

Formative Research

The terms "formative evaluation" and "summative evaluation" originated with Michael Scriven (1967), although the underlying concepts predate these particular labels which are now in common use. Formative research/evaluation deals in general with product or program improvement; it functions to provide evaluative feedback to the production staff while there is still time to incorporate this information and change the program. Formative research applies to the formation or formative stages of a program or product. Summative research, on the other hand, reports only after the program (or subdivision thereof) is completed; it evaluates, after the fact, the extent to which the program (or subdivision thereof) fulfilled its objectives. Stake (1967) added precision to Scriven's terms by contrasting the criteria to which formative and summative evaluations are oriented. Formative evaluation is oriented to "developer-author-publisher criteria and standards." (In the ETV setting, we would probably substitute writer, producer, director.) Summative evaluation, on the other hand, is oriented to "consumer-administrator-teacher criteria and standards." (In ETV we would probably add "sponsoring agency.")

Formative research, like all administrative research, should be an aid in decision-making. The various categories of administrative

research being discussed reflect various categories of decisions that administrators must make. In the ETV formative research category, the decisions deal mostly with message design. The major consumer of formative research is the production staff. The needs of production dictate the agenda for formative research. The criterion by which formative research is judged is the extent to which it was useful to the production department in improving the program. Producers will not set out purposefully to make programs that are boring or incomprehensible, but, for a variety of reasons, such misfortunes do occur. Part of the problem may be a reliance on person A's opinion of what person B will enjoy, understand, pay attention to, learn from, or be persuaded by; in other words, reliance on intuition or advice as a substitute for actual target audience reaction. In evaluative research, there is ultimately no substitute for the reactions of the target audience itself. In program improvement (as opposed to performing an "autopsy" after the resources have already been spent), there is also no substitute for getting this reaction in time to take corrective action. Rapid, sustained feedback to the producers, based on representative audience reaction, is probably the major contribution that formative research can make to the administrative system.

If the purpose of the production staff working with formative research data is to improve the program, and if the major program elements under the control of the production staff are the stimulus elements (i.e., the style and content of the TV program), it follows that formative research in ETV should have stimulus implications. Perhaps it seems strange to make a special point of being relevant to production needs, but, historically, ETV research has not been very

sensitive to the producer's needs, and producers have, for the most part, ignored research findings. However, experience has shown (see, for example, Palmer and Connell, 1970) that production staffs will use research data when it is relevant to their day-to-day problems and decisions, which is precisely the function of formative research. No one who has given serious thought to the communication process would argue that stimulus control equates with control over effects, because this would ignore what the audience contributes to communication effects. The stimulus (the program), however, is the one thing that can be controlled directly by the ETV staff, and considerable attention should therefore be given to it.

This production orientation often leads to a research product that is quite different from the basic research described earlier. For example, in order to get research that can be related to the decisions actually made by TV producers and directors, moment-by-moment effects data are frequently gathered. Whereas most people would probably not instinctively think of evaluating the 17th minute of a program in comparison to the 16th minute, this relates precisely to the complex series of decisions made by a TV director. The physical television program ultimately results from a complex, moment-by-moment series of production decisions made by a TV director. To modify and improve a television program is to change some of these decisions. In this context, moment-by-moment formative research data function quite logically to help locate the specific production decisions that need to be changed.

Consider now the issue of what effect(s), when related to stimulus variables, will be most helpful to the TV producer. What kinds of measures will yield the most insight? Perhaps the most obvious

criterion measures (effects measures) would be the terminal program objectives. The task here would be to relate stimulus variables to performance variables derived from the goals for the program. Gropper et al., (1961) conducted a study (one of a series of excellent studies) designed to measure the amount of improvement that could be brought to a televised lesson by pretesting and subsequently modifying the lesson. Two test lessons were recorded and pretested among a representative group of students. The tests were based directly on the instructional objectives. Test results were analyzed item by item alongside playbacks of the recorded instruction in an attempt to discover the precise points of instructional inadequacy and the underlying reasons for the student difficulties. This analysis constituted the recommendation for a program revision, and a new recording was made. With a considerable degree of rigor and control, performance on the original instruction was compared with performance on the revised instruction, demonstrating improvements in performance ranging from 12 to 26 percent as a function of the one revision.

The situation above involved in-school telecasts with relatively homogeneous groups of students in a supervised setting; the students were a "captive" audience. The heterogeneous, anonymous, non-captive audience that must be attracted to view presents yet another formative research problem. Here, the formative researcher may find it necessary to work back from the terminal objectives to the necessary prerequisites of the terminal objectives -- prerequisites such as attention and comprehension. Unlike face-to-face communication, there is no opportunity in television (except in special two-way systems) to detect evidence of audience inattention or lack of comprehension during

the presentation, and thus no opportunity to take immediate corrective action. Therefore, attention and comprehension cannot simply be assumed. The producers need an advance estimate of these factors, based on program pretests among representative audience members.

Unless the program is attended to, there is no point in further exploration of comprehension or achieving program objectives. Children's Television Workshop (CTW) employs a technique called the Distractor Method to determine the "attention profile" of programs or program segments (Reeves, 1970). A child in such a formative research test would have within his field of vision two things to look at: (1) the TV program being tested, or (2) a series of color slides projected on a rear screen, with the slides changing every few seconds. Actual looking behavior (toward the test program or toward the "distracting" slides) is measured on a quantitative scale repeatedly over time, thus yielding the attention profile. Other recent reports that incorporate observational techniques of actual viewing behavior are Bertram (1971) and Ward (1971).

One of the earliest systematic approaches to diagnostic, moment-by-moment analyses of program effects utilized a device called a program analyzer, through which test audience members could, during the actual monitoring of a program, register interest or disinterest (or other reactions) on a continuous basis by means of push buttons. Lazarsfeld, who along with Frank Stanton developed program analyzer procedures, feels that the procedures are still very useful for what in this chapter would be called formative research:

> Perhaps such a device is most useful for finding out reactions to programs not yet on the air. With the proper research design it would provide one way of

> studying the detailed reactions of unsophisticated viewers to sophisticated programming, and can provide clues to the serious problem of raising the level of sophistication. Program analysis has been used primarily to change the content of programs, to edit or delete difficult parts, to avoid the so-called boomerang effect; however, it could as well play a role in helping to locate people's viewing difficulties. If we could learn something about the stumbling blocks less-educated people have when viewing, let us say, a serious play -- and the program analyzer provides a useful technique for getting at this -- we could then experiment to find out whether supplementary aids, like program materials, or an introductory discussion of the play, have any noticeable effect (Lazarsfeld, 1971, pp. 197-198).

Belson (1967) notes the significant gap in education and technical sophistication that typically separates the broadcaster and his audience. This means that the broadcaster's estimate of what the audience can comprehend should be, if not suspect, at least subject to empirical test. As is the case for all formative research, such tests do not necessarily involve complex techniques. Belson, for example, finds great utility in presenting program material (radio or TV) to people brought together specifically for the testing exercise, then asking them specific (written) questions over the material. Correct answers to the questions would indicate comprehension of that part of the program. If the proportion of comprehension is low, the major formative research effort would be to analyze the patterns of miscomprehension so as to get insight into what is needed to improve program comprehensibility. Formative researchers at Children's Television Workshop can obtain comprehension data from individual subjects on a moment-by-moment basis by stopping the (recorded) presentation in mid-program and asking questions targeted to the immediately preceding information (Palmer, 1972).

Formative research on comprehension is not restricted to verbal content; it can also probe the comprehension of production conventions and a variety of nonverbal content. Chu and Schramm (1967, pp. 137-140) discuss one aspect of this area of comprehension: the problem of communicating with verbally and/or visually illiterate people. They cite a variety of studies in which pictorial conventions were not understood (e.g., use of close-ups, showing only part of an object, etc.) and conclude: "We would be taking too much for granted if we assumed that people having different experiences with visual images would see a film or still picture the way we do" (pp. 138-139). The more sophisticated the encoding of meaning by way of symbolism and production techniques, the greater the risk of miscomprehension or lack of comprehension at the decoding end. Theoretical issues dealing with the "language" or the "grammar" of film and television cannot be discussed here, but the interested reader may refer to Wollen (1969); Davis (1960); Spottiswoode (1959); Whitaker (1970); and Pryluck (1968) for various points of view. The point to be made here is that <u>any</u> meaning, intentional or unintentional, verbal or nonverbal, that is encoded in a television program should be a potential candidate for formative research in comprehension.

TV can communicate in a number of ways simultaneously because the nonverbal channels are so unrestrained with this medium. Much nonverbal behavior is the product of informal cultural norms that are not codified formally but are nevertheless faithfully taught, learned, and adhered to within a culture (see, e.g., Hall, 1969). For the same reasons that it is possible for someone in culture A to insult or offend a person in culture B (i.e., violate his uncodified cultural norms)

through the unwitting exercise of nonverbal communication, it is possible for unintended messages to be conveyed nonverbally via TV. The greater the cultural diversity between producer and audience, the greater the potential problem. Again, formative research can be helpful. Programs can be pretested on the target audience for the purpose of finding out what unintended and undesired messages might be contained. Random samples of the target audience would usually not be required here, and not necessarily even large samples. If a norm were truly universal (adhered to without exception) within a cultural category, it follows that even one representative audience member could indicate if a pretest TV program was offensive or subject to misinterpretation. If management has more of a need for general and sustained feedback reactions within a cultural group or subgroup than for the specific pretest data mentioned above, it should be fairly simple to set up voluntary viewing panels from the target audience who could provide evaluative feedback on a sustained basis, including reactions to nonverbal aspects of the programming.

From the discussion and examples above, it should be apparent that formative research in ETV can take many forms. The unifying theme is that such research or evaluative data should be of utility to decision makers in the production-programming area. Formative research should help them improve their product, the TV program. The logic of each situation will dictate whether such formative research feedback must come before a program is aired or whether it can still serve a useful formative research function after being broadcast (as, for example, in the case where evaluative feedback from an early program in a series can affect production decisions for later programs in the series).

Both background research and formative research should work toward giving the ETV program a reasonable chance of achieving its objectives. Research reports in these two categories are usually not of interest to the general public or even to agencies outside the broadcast operation, nor do they specify what, if anything, was gained from the ETV expenditure. There may be temptations, therefore, to place research money in more "visible" categories. All too frequently, critical pre-broadcast decisions are short-changed in terms of resources devoted to research.

Summative Research

Summative research, as indicated in the previous section, assesses the extent to which program objectives were achieved. The sponsor, the administrative staff, the consumer, and frequently the general public are interested in these evaluations. "How well did it do?" is an inherently interesting question. Issues of high importance in formative research (e.g., does format X or format Y hold attention better for audience group Z?) may not even be mentioned in summative research, because they belong to a different chain of decisions and decision makers. However, when the station manager asks "should I cancel this program or continue it?"; when the head of a government agency asks "should I fund this TV series for another season or not?"; when a school principal asks "should I commit my school to this TV series for next year or continue as before?"; the need for an overall assessment of effectiveness is obvious.

In some form, summative research must be based on a specification of objectives for the program or series. Belson's comments below

were made in the context of comprehension tests, but they apply with equal relevance to summative research:

> It is necessary right at the start to develop a clear idea about what the program material is supposed to be telling people. It is not unfair to say that at times this can be difficult to ascertain even with the script available for careful study. With lengthy or complex material it may be necessary for the question designer to check with the writer himself in order to ensure that his interpretation of the message is what the broadcaster intended (Belson, 1967, p. 126).

The situation implied in Belson's comments is not ideal, but it may well be typical. The unfortunate implication is that the essential programming and production decisions were made before objectives were specified, or at least before the evaluation team was consulted. The situation where the evaluator must look at a finished program and then infer what the specific objectives must have been, so he can then design an instrument to check performance, has happened all too often.

A better model would be the summative research procedures followed at Children's Television Workshop for the program Sesame Street (Ball and Bogatz, 1970; Bogatz and Ball, 1971). At the beginning, content and production experts were consulted to develop program objectives that met the social need and were appropriate for the television medium. A few illustrative behavioral objectives for the 3-5 year target audience of Sesame Street were:

> 1. Matching of letters: "Given a printed letter the child can select the identical letter from a set of printed letters" (Bogatz and Ball, 1971, Appendix B).
>
> 2. Recognition of numbers between 1 and 20: "Given the verbal label for a numeral the child can select the appropriate numeral from a set of printed numerals" (Bogatz and Ball, 1971, Appendix B).

3. Visual discrimination: "Given a form the child can find its counterpart embedded in a picture or drawing" (Bogatz and Ball, 1971, Appendix B).

From the beginning, the summative research team was working on the test battery which was, in effect, the operational definition of program objectives. The test battery would include, for example, the actual letters for matching, the numbers for recognizing, and the embedded figures for finding. Through a lengthy prebroadcast period of planning (which is ideal, but not typical), the summative research test battery was developed and administered to target audience groups for baseline data. Summative research was not an afterthought, but was an integral part of the process right from the start. There are many examples of summative research on television programs, of course, but the wide distribution of Sesame Street around the world makes it a good source of examples.

Certain differences of emphasis on how to relate to behavioral objectives should be noted, especially as these involve summative research. For a useful introduction to the case in favor of behavioral objectives, see Mager (1962). The general premise on which behavioral objectives are supported is that one must know what he means by success and, through explicit criteria, be able to recognize success or failure when he sees it before success or failure of a program can be determined. In a scientific sense, such recognition implies observable behavior and acceptable measurement procedures. While on its face this seems entirely non-controversial, several problems can emerge. If there is no acceptable measuring instrument for effect X, some might argue that effect X cannot be considered one way or the other. Yet such restrictions can be contrary to common experience or common sense. For

example, it is doubtful that a music instructor would give up his commitment to music appreciation as an instructional objective even if this quality totally eludes acceptable measurement procedures. If objectives A, B, and C are specified and agreed to in advance, it is conceivable that the rigorous devotee of behavioral objectives would ignore phenomenon D as being irrelevant, when in fact it might be quite relevant. For example, suppose the subject matter is chemistry, there are three objectives, and A, B, and C, which involve conducting and explaining three experiments. Suppose phenomenon D is the student's interest in doing further study in chemistry. It is possible to succeed in the three experiments, but in the process squelch any further interest in chemistry. On a common-sense basis, most of us would not label this "success," in spite of the fact that the pre-specified objectives were met entirely. There is no single "correct" position to take here, but the ETV researcher should be sensitive to the strengths and weaknesses of whatever position he does ultimately take. See Stufflebeam, et al. (1971) for a more detailed discussion. If one understands the concept of process (as presented, e.g., in Berlo, 1960), he will understand that any ETV program will probably have multiple effects, only some of which can be caught in any particular measurement effort, and only some of which will have been planned for as program objectives, and only some of which will relate directly to the program sponsor's area of responsibility. It follows that summative research cannot be exhaustive, but it should attempt to provide evidence in the categories that are or should be most relevant to the administrative decision required.

The fact that ideal experimental conditions are rarely available to the administrative researcher/evaluator working with ongoing

processes does not get him "off the hook"; he must still be accountable for the interpretability of his data; he must still answer questions such as the following:

1. What assurance do we have that effect Y was due to X and not due to something else? Why should we interpret your data the way you do?
2. What assurance do we have that the evidence is consistent and stable? Why should we accept your measures as reliable and valid?
3. On what basis can you generalize beyond the specific units you measured?

Experimental logic or some form of an alternative approximation of experimental logic is needed to handle such questions.

Given the need for rigor, such as scientific experimental design provides, and given the near-impossibility (or at least great difficulty) of employing true experimental designs in ETV summative research, what are the alternatives? A workable general principle would be to understand the logic of the true scientific experiment and then to make as close an approximation to that degree of control as data utility, the test situation, and research resources will allow. Such a general policy should help guide the ETV researcher to generate optimal realistic designs in a number of different settings. While there is a great amount of literature on experimental designs, there is considerably less on the approximations so frequently required in ETV summative research. In the search for alternatives, the reader should find the following references helpful: Belson (1967); Blalock (1964); Campbell (1969); Campbell and Stanley (1963); Webb et al. (1966).

Cross Media Studies as Summative Research/Evaluation

In the fifties and sixties there emerged a category of ETV research in which televised instruction was compared to "conventional" instruction. These are sometimes called the "TV vs. face-to-face" studies, and they number several hundred. Summary references to these studies can be found in Lumsdaine and May (1965); Reid and MacLennan (1967); and Chu and Schramm (1967). In terms of the research categories being used in this chapter (basic research, plus four categories of administrative research), many of the "TV vs. face-to-face" studies would have to be classified as summative research. The "TV vs. face-to-face" studies have been repeatedly and severely criticized on both methodological and conceptual grounds (see, e.g., MacLean, 1962; Stickell, 1963; Kittross, 1967; Mielke, 1968, 1971), and indeed, such cross-media comparisons are now rare.

To judge the adequacy of a methodology is to check on such features as the research design employed, the degree to which assumptions of statistical tests were met, and the like. Although admittedly difficult to apply in on-going processes such as schoolroom classes, high rigor is not difficult to recognize, in the sense that it is defined by rather widely accepted criteria. On methodological grounds alone, Stickell (1963) found only 10 of 250 comparisons of "TV vs. face-to-face" to be "interpretable."

It is somewhat more difficult to judge the adequacy of the question asked in the first place. Basic research seeks to test a rationale or theory. Good basic research questions are formed so as to conceptualize clearly a reason for expecting one variable (or set of variables) to have a demonstrable influence or effect on another

variable (or set of variables). The "TV vs. face-to-face" studies have been impressively lacking in such rationales, and they have made little if any contribution to basic theory.

Exploratory research, which is pre-theoretic, seeks to discover stable relationships among variables so that rationales for these relationships can subsequently be derived and tested. Good exploratory research questions are therefore questions that have a high or reasonable probability of payoff, i.e., of finding stable relationships that can be explained with a scientific theory. For example, consider the scientific study of weather. For exploratory research on the weather to lead to general theories and principles of weather phenomena, at least two things must happen: (1) theoretically relevant factors about the weather must be observed in the first place; and (2) these observations must be made systematically and with enough precision to allow patterns and relationships to become noticeable. The same reasoning holds for ETV-related phenomena. Both lack of precision and lack of theoretic potential have plagued the "TV vs. face-to-face" studies, and little in the way of new or modified theory has emerged as their result.

Administrative research serves administrative requirements rather than theoretic requirements (although these two requirements can sometimes be found to be closely associated). The good administrative research question is one that serves well the needs of the administrative decision maker. Even here, the utility of the "TV vs. face-to-face" studies is lacking. They serve no obvious background, formative, or policy research function, and their summative research function is severely limited by some of the specific problems discussed below. These historical problems with "TV vs. face-to-face" studies illustrate

pitfalls in rationale and procedure that can be repeated in new settings of summative research in ETV.

A. "Is TV 'as good as' face-to-face?" is a poorly conceptualized question. Possible reasons for expecting TV to be "better" (e.g., more released time for lesson preparation; ability to "show directly" rather than "talk about"; distributing the best instruction to all students; etc.) or possible reasons for expecting TV to be "worse" (e.g., inability of students to question the television teacher directly; inability to adjust lesson pacing to the needs of individual students; etc.) might be useful factors to study, because they deal directly with characteristics, correlates, or consequences of the television medium or the television system. To ask the unrefined comparative question, however, is to ask whether unknown package A is better or worse than unknown package B; no matter what the "answer" is, it will be very difficult to understand, explain, utilize, or generalize.

B. The more the two treatment conditions are matched, the less realistic either condition. The more the TV and face-to-face groups are actually matched on all factors except the type of mediation, the less freedom there is for either form of instruction to do what it can do best. Conversely, the more each form or system of instruction exploits its unique potential, the more difficult or impossible the two become to compare and analyze scientifically. Conceptually, the problem of scientific assessment becomes amplified if TV is but one component of an innovative system, as is the case, for example, in Great Britain's Open University (Smith, 1972). Potentially, this introduces a double-edged problem of (a) not being able to detect an effect unless the entire system is activated; and (b) then not being able to attribute

effects to specific components in the system, such as television. The final report of the National Project for the Improvement of Televised Instruction, 1965-1968, notes:

> . . . in any complex situation, if only one variable is changed, no significant difference results. This is why the overall improvement in American education, after ten to fifteen years of "spot" innovations, has been relatively small (Toward a Significant Difference, undated report, p. 36).

The other side of the dilemma is articulated by Guttentag (1971) who writes:

> In contrast to an experimental study, a researcher evaluates the effect of a programme . . . not the effect of a variable. Programmes, however, consist of a multitude of bits and pieces. They are the inverse of the carefully defined and manipulated single variables of the experimental paradigm (p. 76).

C. Even when each method is free to exploit what it does best, the selection of the "effects" to be measured (the dependent variables) frequently cannot reflect differences that may well be present. In part, this problem relates to the problems of specifying behavioral objectives discussed earlier. Just because an effect cannot currently be specified and measured does not mean that it is irrelevant. Just because an effect was not measured does not mean that the effect did not exist. On measures of cognitive learning, one will typically find no statistically significant difference (NSD) between the TV and face-to-face groups. Other studies also find NSD with home study groups and other forms of instruction. One could either argue that (a) the measures of cognitive learning (typically paper-pencil measures of factual recall) tap an impressively robust quality in a variety of settings; or (b) such measures are simply insensitive to other differences that do exist. Let the

researcher examine such an instrument designed to test cognitive learning and ask himself if there is any good reason to expect a TV group, a face-to-face group, a radio group, a correspondence study group, etc., to have a marked advantage in completing such a test successfully. If the answer is "no," then that particular instrument is probably not well suited to detect possible differences among various methods of instruction, such as television.

What might some of the elusive differences be -- i.e., effects that would probably elude traditional tests of cognitive achievement? No special case is made for this particular illustrative list of effects, except the argument (a) that these factors would not be included in a typical "learning" assessment; (b) that such effects could be of critical importance to the success of an ETV program; and (c) that there are situations where TV could conceivably hold a distinct advantage on such factors. Consider these possible effects:

1. ability to relate and apply the information to one's own life (Given a dramatic TV presentation, perhaps a student could identify with a TV character and be greatly affected on this factor.)

2. credibility of the information (Some studies [e.g., Roper, 1969] indicate that differing credibility levels are attributed to various media.)

3. impact on attitudes and values (In a communication with persuasive intent it is usually possible to "learn" the various arguments presented without changing opinion.)

4. amount of status or prestige attributed to the instructor (A frequent phenomenon is that prestige is associated with

appearance on TV.)

5. amount of incidental learning that took place (The combined audio and video channels of television typically provide a wealth of information considered incidental or unrelated to the major thrust of the lesson, at least as defined by teachers. As anyone knows who has spoken with young children who have been exposed to a substantial amount of entertainment TV programming, however, such incidental learning can be significant.)

The list could be greatly extended, but these few entries can serve to illustrate non-obvious effects that could escape notice if the only opportunity to register an effect was a cognitive learning measure. Even if each media system exploits what it does best, this may remain unknown and unappreciated if the criterion tests are insensitive to unique contributions of each system.

D. In "TV vs. face-to-face" studies, there is an implicit assumption that the criterion for TV to "measure up to" is traditional instruction. Such a comparison does not ask whether the traditional instruction is satisfactory in the first place, nor does such a comparison serve any diagnostic function for either group.

In the more general field of educational evaluation, group comparisons have been similarly criticized by Cronbach (1963). However, Scriven (1967) still argues that when the administrator wants to know if method A is "better" than method B, the direct comparison is the best way to find out. Scriven argues that demonstrated superiority of version A can allow a series of useful administrative decisions even if it is unknown why version A was superior and even if A was compared to B as an expedient because no absolute scale of evaluation for A was

available. For the ETV situation, the weight of experience in evaluative research would seem to be against reviving the comparative study of TV vs. face-to-face, but this would not rule out the comparative study of TV version A vs. TV version B (e.g., Gropper and Lumsdaine, 1961) which would more nearly satisfy the criteria of experimental design.

Policy Research

> . . .if television is changing its viewers, it is highly desirable that the public and those who control television should be aware of the nature and the extent of these changes and of precisely how they come about. Such information seems to the writer to be essential in any attempt to control the television medium in the interests of society (Belson, 1967, p. 227).

The four categories of ETV administrative research discussed in this chapter (background research, formative, summative, and policy research) correspond roughly and imprecisely with the four categories of TV research developed by Belson (1967): planning research, comprehensibility research, effects of specific programs, and social impact research. Of the four, the fit is probably best between our category of "policy research" and Belson's "social impact" category. If summative research can be conceptualized as feeding into administrative "recycling" decision-making (Stufflebeam, et al., 1971), then policy research would be providing input to recycling decisions at a more general societal level.

Many policy issues involving television become policy issues primarily because of the magnitude of TV-related phenomena. The magnitude of investment in hardware, the magnitude of TV's reach, the magnitude of the amount of time devoted to watching TV, the magnitude of the

opportunities if TV is used creatively and wisely -- these magnitudes raise serious policy issues that transcend the broadcasting organization. Indeed, such policy issues extend to the highest levels of government and other societal institutions.

Sometimes, in response to articulated problems and needs, policy research is commissioned by policy-responsible groups such as government. The U.S. Government, for example, has in the last few years received reports from such commissions as:

1. The Commission on Instructional Technology (1970);
2. The Commission on Obscenity and Pornography (1970);
3. The Surgeon General's Report on Television and Social Behavior (1971).

In varying degrees, such reports, which tend to be combinations of research reports and expert testimony, have influenced policy in mass communication, although the level of influence is rarely as great as that advocated by social scientists.

Sustained monitoring/evaluation of media performance (see, e.g., Lasswell, 1972) could be an example of policy research. Non-obvious effects of mass communication (e.g., effects on physical well-being, eating habits, church attendance, amount of conversation, etc.) frequently have policy implications much broader than the domain of any particular ETV program. Such policy-relevant effects would, in Lazarsfeld's (1971) view, have long-range repercussions. A long-range repercussion means an effect that is not only sustained over time, but one that "also spreads out into the social field: an effect may act not only on one person but on several and, finally, on entire institutions as, for example, on education" (p. 192).

In the general and idealized case, the advocate for policy research would be an advocate for basing policy on fact, for empirically testing many premises and assumptions of policy statements. If grounded in scientific method, such a policy research advocate would already have analytic experience with the concepts of multiple causation and multiple effects, so he would avoid oversimplification of policy issues. The policy research advocate would be willing to change policies on the basis of new evidence in approximately the same way that scientists would be willing to revise theories on the basis of new evidence.

Final Notes

No matter how frustrating the requirements of experimental rigor may be to the administrative researcher, an experimental type of question will only yield to an experimental type of procedure and logic (i.e., a true experiment or some approximation to an experiment). If one would reject the experimental logic, he would, to his impoverishment, also have to reject the type of questions which only experimental logic can deal with. On the other hand, no matter how much the general rigor or generalizability of evaluation studies may be criticized by the basic researcher, experimental methodology simply cannot cope realistically with many administrative questions and problems. If one would reject all questions that are not derived from a general theory and that will yield to an experimental design, he would be turning many administrative decisions over to tradition and guesswork. The administrative decisions must be made, with or without research data input.

A major judgmental issue, then, is to decide what question it is that must be answered. Once this is determined, methodological choices are at least partially delimited. Superiority or inferiority of a research method cannot be established as an inherent quality, but it can be established in terms of performance in answering the question. Given this view, debate on the relative merits of "basic" vs. "applied" research per se, or "formative" vs. "summative" evaluations per se, will be recognized as generally futile. Real world problems will pose a variety of questions, requiring a variety of research methods in response.

As a means of tying together the considerable variety of analyses and arguments on ETV research presented in this chapter, consider a hypothetical case of a decision to produce a TV series on subject X. From a very large set of possible questions, a few representative questions have been conjectured here to illustrate both the relevance and interdependence of various research inputs. Brief comment will follow each question or set of questions.

Q. 1-a: WHAT IS KNOWN ABOUT SUBJECT X?

Q. 1-b: WHY DO PEOPLE HAVE THE BELIEFS AND ATTITUDES THEY DO ABOUT X?

Q. 1-c: HOW RESISTANT TO CHANGE SHOULD WE EXPECT THESE BELIEFS AND ATTITUDES TO BE?

These three illustrative questions reflect the need for a theoretic grasp of the subject matter as well as a theoretic grasp of the relationship of the audience to the subject matter. Basic research may not be the typical mission of the ETV researcher assigned to a broadcast operation, but basic research and scientific theories founded on such

research can be very useful for practical decision-making in ETV. Suppose "subject X" is malaria, and ETV is being considered as an instrument to wage a nation-wide campaign against malaria. Imagine the difficulty or even futility of trying to implement a TV campaign in the absence of known or assumed answers to these three questions. The real issue is probably whether the operating assumptions of the ETV programmer are or are not to be supported by scientific theory and basic research.

Q. 2: WHAT ARE THE COMMUNICATION HABITS OF THOSE MOST IN NEED OF KNOWLEDGE ABOUT X?

This kind of background research can be very useful to the ETV programmer. At a common sense level, such background research helps one to avoid certain pitfalls; e.g.:

A) If the target audience cannot read, don't depend on printed program supplements;

B) If the target audience is in bed at 10 o'clock, don't schedule your program for 10 o'clock;

C) If the target audience is most fluent in dialect B, don't broadcast your program in dialect A.

In a positive vein, knowledge of communication habits can be critical in planning general strategy, as exemplified very well in some of the work done by Harold Mendelsohn (1968; 1969; 1971). If the vast majority of the target audience already is exposed to television broadcasts, that is an important and positive factor in favor of using television for the campaign or social development program on subject X. If the target audience reliably seeks advice on subject X from a certain person or a certain institution, it implies to the TV strategy

planner that the help of this certain person or institution should be enlisted from the beginning. The general thesis here is that the target audience never exists in a vacuum; it always operates in a complex social system, and the objective is to work as compatibly as possible with, rather than against, the audience's existing social system and habit patterns.

Q. 3: CAN THE NATURE AND POTENTIAL OF THE TV MEDIUM BE USED TO ANY SPECIAL ADVANTAGE FOR INFORMING THE TARGET AUDIENCE ABOUT X?

This question could be overlooked by the planner who conceptualized TV primarily as an efficient distribution system. One obvious approach to the potential of the medium lies in an examination of production capacities. Of the dozens of things that a creative producer can do with TV, which can be used with greatest artistic or pedagogical advantage? Slow motion, rapid motion, time-lapse photography, instant replay, and multiple images are just a few of the tools at the disposal of the TV (or film) producer/director that are not commonly available in other media. Planning with sensitivity to the medium will include conscious consideration of the unique message design features made possible by the medium. For an analysis of media characteristics and their implications for message design, see Bretz (1969). Perhaps a less obvious approach to the nature of the medium and its potential advantages for programming about subject X would be through a sociological/psychological analysis of the TV reception environment. Sometimes a weakness from one point of view can be a strength from another. For example, Palmer (1969) notes that in the Sesame Street series a potential disadvantage (being unable to provide immediate corrective feedback) was

at least partially offset by a corollary advantage (the programs were not threatening; poor performance was not ridiculed or punished). When viewing is done in small, private groups, a person may be willing to receive instruction or other communication on topics that he or she would be unwilling or too embarrassed to receive in a more public setting. For example, topics related to personal hygiene, sex education, and venereal disease might be able to exploit the nature of television rather than be handicapped by it.

Q. 4-a: WHAT ARE REALISTIC OBJECTIVES FOR A TV SERIES ON X?

Q. 4-b: HOW CAN THESE BE TRANSLATED INTO PROGRAMMING CONCEPTS?

The problem is not only to determine the objectives and curriculum, which requires knowledge of the subject matter and its relationship with the intended learner, but to determine realistic objectives and curriculum for <u>television programs</u>. This adds the requirement briefly discussed above in Q. 3: knowledge of the nature and potential of the television medium. For example, this requirement was recognized in the planning seminars when the children's program <u>Sesame Street</u> was being conceptualized, as reported by Lesser (1972):

> Early in our discussions we did not allow ourselves to be constrained by what the production staff thought feasible in the production of a show. But as we came to setting priorities this became a primary concern. The creative producers and writers reminded us that our task was the concrete and practical one of constructing a set of educational goals that could be understood clearly and converted into actual television programming. They were especially insistent when other seminar participants used esoteric jargon. On these occasions, the staff resembled a Greek chorus, intoning repeatedly, "What do you mean by that? What do you mean by that?" (p. 234).

Questions 3 and 4 represent types of questions and issues that need to be grappled with in the administrative process; they do not necessarily require the execution of original research, although original research would be entirely appropriate. In the category scheme used in this chapter, such research would probably be considered a very early stage of formative research.

Q. 5: OF FIVE AVAILABLE TALENTS, WHICH ONE WILL THE TARGET AUDIENCE FIND MOST INTERESTING AND CREDIBLE?

This question is intended to represent a host of situations where the ETV administrator must make a selection from a finite set of realistic options. Suppose a station is trying out five newscasters for a single position. Formative research can provide data input to be considered alongside other inputs in the decision-making process (that is, the decision on which newscaster to hire will undoubtedly be based on more factors than interest and credibility ratings). To the theorist, this question would have little interest because the variable represented by the five options cannot be interpreted in a theoretically meaningful sense. To an experimentalist, this question would also have little interest because the five choices do not constitute a manipulable variable; the question asked is not an experimental kind of question, such as what is the effect of X on Y. The formative researcher would probably agree also that the question is not profound, but he would see that a decision had to be made, and that he could contribute to the rationality of that decision by providing evidence. The question seems to call for some form of survey methodology. The formative researcher now faces a series of sub-decisions: how much time and money can he devote to answering the question; and, within those constraints, how

much time and money should he devote to answering the question? If the hiring decision must be made tomorrow, an elaborate survey is, of course, out of the question. If there is time, the survey may be ruled out by lack of budget. If time and budget, it may be ruled out or compromised as a value judgment on the relative importance of this expenditure versus other anticipated expenditures for the formative research budget later on. Almost by definition, the formative researcher deals continuously in compromise.

Q. 6: BEFORE THIS PARTICULAR PROGRAM ON X IS RELEASED FOR BROADCAST, WHAT REASONS DO I HAVE TO SUPPOSE THAT THE TARGET AUDIENCE WILL FIND IT APPEALING, INOFFENSIVE, COMPREHENSIBLE, MEMORABLE, AND EFFECTIVE?

A number of questions have been grouped into the single question above because it represents a broad and important area of formative research: program pretesting and improvement. Such formative research is empirical in the sense that it seeks to test assumptions. If a writer feels that a certain critical scene will, as desired, be very funny, the formative researcher can try it out on a test audience and see if anybody laughs. Such formative research is pragmatic in the sense that recommendations will be based on what has or has not worked successfully, as determined from all available sources, even if this is not yet fully understood in a theoretical sense. Such pretesting can be over the specific terminal objectives for the program or series; it can be directed toward a series of prerequisite or instrumental objectives such as attention and comprehension; and it can serve as a warning system for other problems totally unanticipated, such as an

unintentional but offensive violation of cultural norms.

Q. 7-a: DID THE TARGET AUDIENCE REACH THE DESIRED LEVEL OF PERFORMANCE OR KNOWLEDGE ABOUT X BY THE END OF THE SERIES?

Q. 7-b: WHAT WOULD THE TARGET AUDIENCE HAVE KNOWN ABOUT X IF THE TV SERIES HAD NOT BEEN SHOWN?

Answers to these questions are somewhat like the final score in a game: they indicate whether you won or lost. Achieving program objectives is the goal toward which all the theorizing, background research, and pretesting have been directed. The simplistic athletic analogy breaks down rapidly as actual complexities are considered, such as the variety of objectives and the variety of measurement approaches that might be applied to a program. For example, the administrator may know that he succeeded on criterion A; know that he failed on criterion B; be forced to rely on expert opinion on how he did on criterion C; be unable to afford to find out how he actually did on criterion D; and have no idea how he did on criterion E. If the program is expensive and/or controversial, there will probably be critics actively searching for negative evidence and finding fault with the summative research report.

The "Four R's" for summative research might be called Rigor, Relevance, Realism and Responsibility. These values do not necessarily call forth the same kinds of activity. (Although no special questions were generated to illustrate policy research, the "Four R's" apply equally well to policy research.) Rigor seeks unequivocal results. Relevance demands that the answers be useful to actual decision needs. Realism recognizes that many beneficial reforms and innovations depend on political values as well as scientific values. Responsibility tries

to find out what is happening and what could happen, so that a thoroughly considered value system can direct the choice of what should and will happen. A sensitivity to all four of these values is reflected in the following statement by Campbell (1969):

> What is . . . essential is that the social scientist research advisor understand the political realities of the situation, and that he aid by helping create a public demand for hard-headed evaluation, by contributing to those political inventions that reduce the liability of honest evaluation, and by educating future administrators to the problems and possibilities (p. 409).

REFERENCES

Ball, S. and Bogatz, G. A. The First Year of Sesame Street: An Evaluation. Princeton, New Jersey: Educational Testing Service, 1970.

Belson, W. A. The Impact of Television: Methods and Findings in Program Research. Hamden, Connecticut: Archon Books, 1967.

Berlo, D. K. The Process of Communication. New York: Holt, Rinehart and Winston, 1960.

Bertram, C. L. Evaluation Report: Early Childhood Education Program, 1969-1970 Field Test. Summary Report. Charleston, West Virginia: Appalachia Educational Laboratory, May 1971. Also see Analysis of Children's Reactions to AEL's Preschool Television Program. Technical Report No. 9. Charleston, West Virginia: Appalachia Educational Laboratory.

Blalock, H. M., Jr. Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press, 1964.

Bodenave, J. D. Orientacion Desarrollista en la Comunicacion Colectiva. San Jose, Costa Rica, Servicio Interamericano, Instituto Interamerica de Ciencias Agricolas de la OEA, 1965.

Bogatz, G. A. and Ball, S. The Second Year of Sesame Street: A Continuing Evaluation. Princeton, New Jersey: Educational Testing Service, 1971.

Bretz, R. Communication Media: Properties and Uses. Santa Monica, California: Rand Corporation, 1969.

Campbell, D. T. "Factors Relevant to the Validity of Experiments in Social Settings," Psychological Bulletin, 1957, 54, 297-312.

Campbell, D. T. "Reforms as Experiments," American Psychologist, April 1969, 24 (4), 409-429.

Campbell, D. T. and Stanley, J. C. "Experimental and Quasi-Experimental Designs for Research on Teaching." In N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Chu, G. C. and Schramm, W. Learning from Television: What the Research Says. Stanford, California: Stanford University, Institute for Communication Research, 1967.

Connell, D. D. and Palmer, E. L. Cooperation Between Broadcasters and Researchers. Leicester, England: International Seminar on Broadcaster/Researcher Cooperation in Mass Communication Research, 1970.

Cronbach, L. J. "Course Improvement Through Evaluation," Teachers College Record, 1963, 64, 672-683.

Davis, D. The Grammar of Television Production. New York: Macmillan, 1960.

Gropper, G. L. and Lumsdaine, A. A. Studies in Televised Instruction: A Summary Report. Pittsburgh: American Institute for Research, June 1961.

Gropper, G. L., Lumsdaine, A. A., and Shipman, V. Improvement of Televised Instruction Based on Student Responses to Achievement Tests. Studies in Televised Instruction, Report No. 1. Pittsburgh: Metropolitan Pittsburgh Educational Television Stations WQED-WQEX and American Institute for Research, March 1961.

Guba, E. G. "The Failure of Educational Evaluation," Educational Technology, May 1969, 9 (5), 29-39.

Guttentag, M. "Models and Methods in Evaluation Research," Journal for the Theory of Social Behavior, 1971, 1 (1), 75-95.

Hall, E. T. The Silent Language. Greenwich, Connecticut: Fawcett, 1959.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1965.

Kittross, J. M. "Meaningful Research in ETV," in A. E. Koenig and R. B. Hill (eds.), The Farther Vision: Educational Television Today. Madison, Wisconsin: The University of Wisconsin Press, 1967. pp. 209-245.

Lasswell, H. D. "Communications Research and Public Policy," Public Opinion Quarterly, 1972, 36 (3), 301-310.

Lazarsfeld, P. F. "Trends in Broadcasting Research in the United States," in H. Eguchi and H. Ichinohe (eds.), International Studies of Broadcasting. Tokyo: Japan Broadcasting Corporation, 1971.

Lesser, G. S. "Learning, Teaching, and Television Production for Children: The Experience of Sesame Street," Harvard Educational Review, 1972, 42 (2), 232-272.

Lumsdaine, A. A. and May, M. A. "Mass Communications and Educational Media," Annual Review of Psychology, 1965, 16, 475-534.

MacLean, M. S., Jr. "Critical Analysis of 12 Recent Title VII Research Reports," AV Communication Review, 1962, 10 (3), A-102 - A-114.

Mager, R. F. Preparing Instructional Objectives. Belmont, California: Fearon, 1962.

Mendelsohn, H. Operation Gap-Stop: A Study of the Application of Communication Techniques in Reaching the Unreachable Poor. University of Denver, Communication Arts Center, February, 1968.

Mendelsohn, H. "What to Say to Whom in Social Amelioration Programming," Educational Broadcasting Review, 1969, 3 (6), 19-26.

Mendelsohn, H. "Cancion de la Raza Evaluated," Educational Broadcasting Review, 1971, 5 (5), 45-53.

Mielke, K. W. "Asking the Right ETV Research Questions," Educational Broadcasting Review, 1968, 2 (2), 54-61.

Mielke, K. W. "Evaluation of Learning from Televised Instruction," in R. C. Burke (ed.), Instructional Television: Bold New Venture, Bloomington, Indiana: Indiana University Press, 1971, pp. 97-113.

Mielke, K. W. "Renewing the Link Between Communications and Educational Technology," AV Communication Review, 1972, 20 (4), 357-399.

Palmer, E. L. "Research at the Children's Television Workshop," Educational Broadcasting Review, 1969, 3 (5), 43-48.

Palmer, E. L. "Formative Research in the Production of Television for Children." Paper presented at the International Symposium on Communication: Technology, Impact, and Policy. Annenberg School of Communication, University of Pennsylvania, March 23-25, 1972.

Pryluck, C. "Structural Analysis of Motion Pictures as a Symbol System," AV Communication Review, 1968, 16, 372-402.

Reeves, B. F. The First Year of Sesame Street: The Formative Research. Final Report. Vol. 2. New York: Children's Television Workshop, December, 1970.

Reid, J. C. and MacLennan, D. W. Research in Instructional Television and Film. Washington, D.C.: U.S. Government Printing Office, 1967.

A Report by the Commission on Instructional Technology, in S. G. Tickton (ed.), To Improve Learning: An Evaluation of Instructional Technology. New York: R. R. Bowker Co., 1970.

The Report of the Commission of Obscenity and Pornography. New York: Bantam Books, 1970.

Roper, B. W. A Ten-Year View of Public Attitudes Toward Television and Other Mass Media: 1959-1968. New York: Television Information Office, 1969.

Salomon, G. "What Does it do to Johnny?" in G. Salomon and R. E. Snow (eds), Commentaries on Research in Instructional Media. Indiana University: Viewpoints, 1970, 46 (5), 33-62.

Salomon, G. "Can We Affect Cognitive Skills Through Visual Media? An Hypothesis and Initial Findings," AV Communication Review, 1972, 20 (4), 401-422.

Salomon, G. and Snow, R. E. "The Specification of Film Attributes for Psychological and Educational Research Purposes," AV Communication Review, 1968, 16, 225-244.

Scriven, M. S. "The Methodology of Evaluation," AERA Monograph Series on Evaluation, No. 1. Chicago: Rand McNally, 1967, pp. 39-83.

Smith, P. J. "Britain's Open University: Everyman's Classroom," Saturday Review, 1972, April 29, 40-42, 47-50.

Spottiswoode, R. A Grammar of the Film. Berkeley: University of California Press, 1959.

Stake, R. E. "The Countenance of Educational Evaluation," Teachers College Record, 1967, 68 (7), 523-540.

Stickell, D. W. A Critical Review of the Methodology and Results of Research Comparing Television and Face-to-Face Instruction. Ed.D. thesis, The Pennsylvania State University, 1963.

Stufflebeam, D. L. "The Use and Abuse of Evaluation in Title III," Theory Into Practice, 1967, 6 (3), 126-133.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., and Provus, M. M. Educational Evaluation and Decision-Making. Bloomington, Indiana: Phi Delta Kappa, 1971.

Television and Growing Up: The Impact of Televised Violence. Report to the Surgeon General from the Surgeon General's Scientific Advisory Committee on Television and Social Behavior. Rockville, Maryland: National Institute of Mental Health, 1971.

Toward A Significant Difference: Final Report of the National Project for the Improvement of Televised Instruction, 1965-1968. Washington, D.C.: National Association of Educational Broadcasters, undated report.

Ward, S. Effects of Television Advertising on Children and Adolescents. Cambridge, Massachusetts: Marketing Science Institute, 1971.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

Whitaker, R. The Language of Film. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Wollen, P. Signs and Meaning in the Cinema. Bloomington, Indiana: Indiana University Press, 1969.